


How to Link Your Institution's Collections to Wikidata? : a short manual to a semi-automatic way of using the “archives at” property (P485)

Other Publication**Author(s):**

Bolliger, Stephanie; Brüderlin, Brigitte; [Gasser, Michael](#) ; Lyskawa, Julia; Maier, Petra; Schmitt, Lothar

Publication date:

2020-01

Permanent link:

<https://doi.org/10.3929/ethz-b-000393724>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

How to Link Your Institution's Collections to Wikidata?

A short manual to a semi-automatic way of using the “archives at” property (P485)

Published in January 2020

Table of contents

Goal of a co-operative project.....	1
Documentation.....	2
Workflow overview	2
Metadata extraction.....	2
Identification of existing items in Wikidata and Wikidata import with OpenRefine	3
Tips and tricks.....	13
Additional resources (last retrieved: January 16 2020)	13

Goal of a co-operative project

The aim of the project members of the three Swiss cultural heritage institutions ETH Library, Zurich Central Library and Swiss National Library, Berne, carried out together between June and November 2019 was to interlink existing (online) archival finding aids with their respective Wikidata items of persons. In Wikidata the property “archives at” ([P485](#)) is used to specify the institution(s) holding the subject's archives or private collections.

This example from Leopold Ružička ([Q122996](#)) illustrates the resulting, very simple structure of the use of P485:

The screenshot shows the Wikidata edit interface for the property 'archives at'. A reference is added for 'ETH Zurich University Archives' with the inventory number 'CH-001807-7:Hs 1191'. The reference URL is 'http://archivdatenbank-online.ethz.ch/hsa/#/content/149eeab3aac44d31ab55094d61ce820f'. The reference was retrieved on '22 October 2019'. Annotations with arrows point to the 'Institution (existing Wikidata item)' label, the 'reference URL' field, and the 'Link to the online finding aid' text.

archives at	ETH Zurich University Archives	edit
inventory number	CH-001807-7:Hs 1191	[[ISIL code]:[inventory number] (string)
reference URL	http://archivdatenbank-online.ethz.ch/hsa/#/content/149eeab3aac44d31ab55094d61ce820f	Link to the online finding aid
retrieved	22 October 2019	

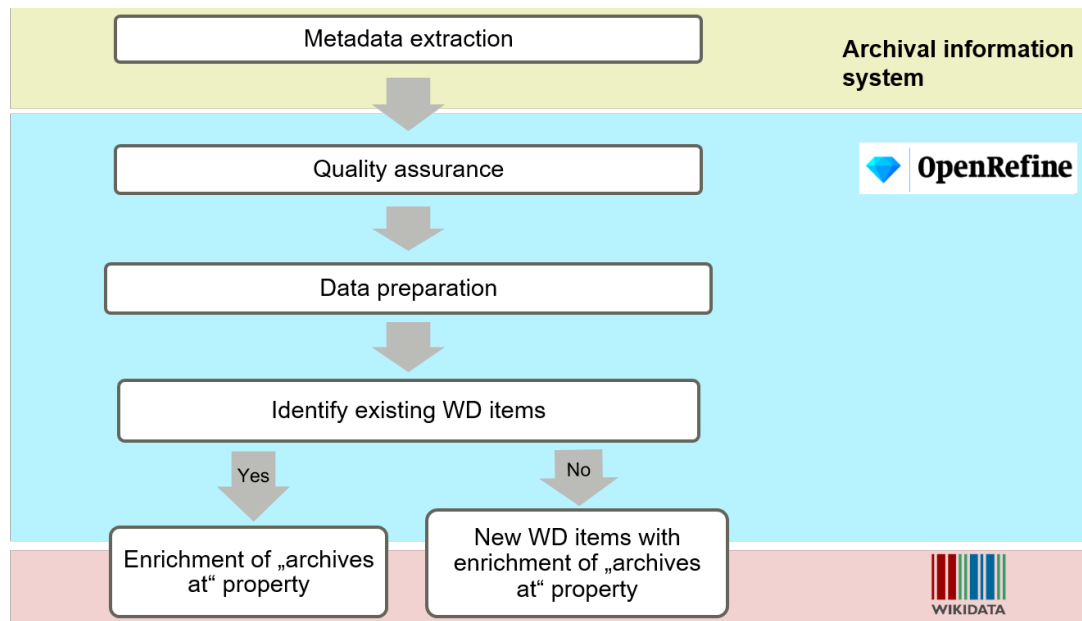
+ add reference

If only a limited number of collections (up to around 100) are to be linked to Wikidata, this is best done manually. In our project, however, each institution set out to link several hundred finding aids. This is why we established a semi-automatic way to achieve our project goal. Both for the case that

existing Wikidata items are just enriched by using the property "archives at", and for the case that new Wikidata items have to be created first. We tried to document this process in the present short manual to pave the way for other institutions to increase the visibility of their holdings in Wikidata and the world of linked open data that comes with it.

Documentation

Workflow overview



Metadata extraction

Of course, metadata extraction heavily depends on the source system in which the descriptive information is stored, the way the metadata is structured, and the interfaces available. In our project, initial ISAD(G) structured metadata was extracted from three different archival information systems as CSV or Excel files.

We focused on the following properties for each private collection:

- surname
- first name
- year of birth (if available)
- year of death (if available)
- descriptive information (if available)
- archival reference number
- link to the respective online finding aid

If the data are semi-structured (e.g. a title like "Private papers of XY (1910–1985)") Excel formula or regular expressions can help to extract and prepare the data.

Data quality assurance options

In order to improve the data quality, we would like to give some general recommendations:

- exclude incomplete/insecure data
- OpenRefine:
 - run tests with a limited amount of data
 - data preparation: try how to get the best reconciliation results (e.g. order of the name: given name + family name, year of birth and year of death separated or in 1 field with hyphen)
 - reconciling against Wikidata, double check of the results
 - reconciling against other authority files, e.g GND (=German authority file) to get more data

Identification of existing items in Wikidata and Wikidata import with OpenRefine

OpenRefine: step by step (OpenRefine Versions 2.8, 3.1, and 3.2)

In order to retrieve and prepare the data you want to import to Wikidata we provide you with the following short step by step procedure.

1. Reconciliation against Wikidata

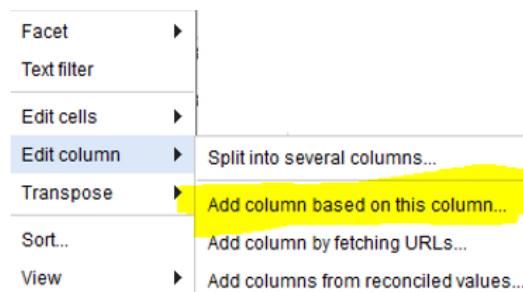
1.1. Data preparation for reconciling against Wikidata

i

Add a column with the names (=Label) in the following structure:

Given name + family name (e.g.: *Julius Alpari*)

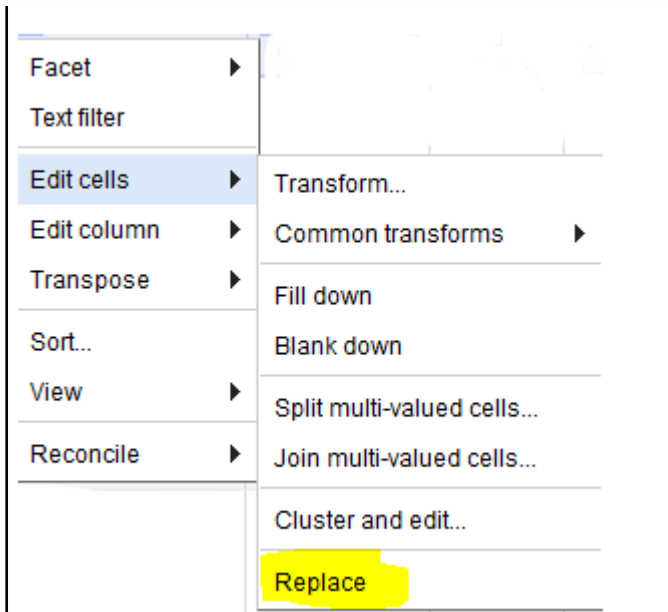
Add an identical separate column to be used for data reconciliation.



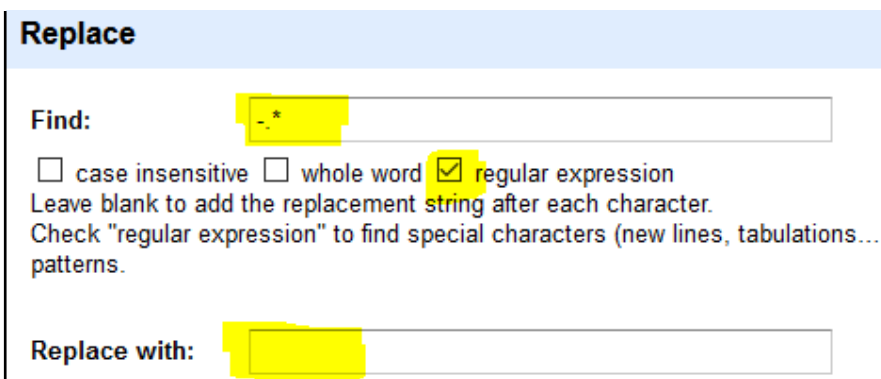
1.2. Reconciliation against Wikidata

“Reconcile” -> “Start reconciling” -> Services: Wikidata

Select “human” (Q5) on the left side and input “date of birth” (P569) and “date of death” (P570) in the right column as additional details for better reconciliation results.

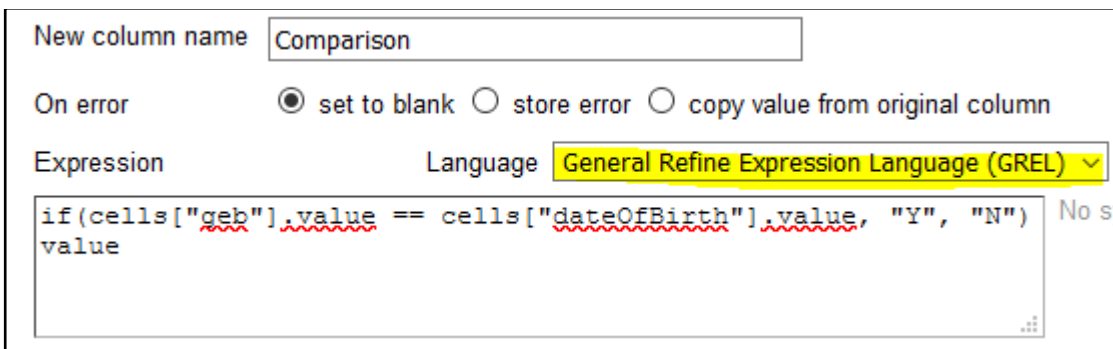


Find: -.* [select the option "regular expression"]
 Replace: [empty field]
 -> this expression matches the first hyphen - and all following characters .*
 The result also has to be formatted as "text".



Comparison of "date of birth" and "date of death" from Wikidata with your original values:
 "Edit column" -> "Add column based on column":

```
if(cells["Column 1"].value == cells["Column 2"].value, "Y", "N")
```



Use the text filter function (filter by "N") to identify those records where the original and the imported values do not match.

i Depending on your data it can be helpful to repeat this step also for the date of death and the name.

1.5. Include Wikidata-ID (= Q-number) of the matched values

"Edit column" (Label) -> "Add column based on column": `cell.recon.match.id`

2. How to split the entire dataset into different subsets?

At this point the entire dataset has to be split up into two datasets (records with existing Wikidata items and records without):

! The easiest way to do this is by using the faceting function -> facet the column with the Q-number: with ID / without ID

If you plan to import your data into Wikidata via OpenRefine (Wikidata extension in OpenRefine "Export to QuickStatements"), you need different QuickStatement schemas. But you can save only one schema per project. Therefore, it is best to continue processing the data in different projects -> Export project

Filter the records with Wikidata Q-numbers (a Wikidata item already exists):

"Text filter" searching for all cells starting with the letter Q. Filter the column with Wikidata Q-numbers by Text Facet. Choose the blank ones for one set. Remove all matching rows.

The screenshot shows the OpenRefine interface. On the left, the 'Facet / Filter' panel is active, showing a text facet on the 'Q.-ID' column. The facet has 329 choices, sorted by name, with counts. The choices include Q76483 (1), Q76815 (1), Q77109 (1), Q813089 (1), Q824840 (1), Q87292 (1), Q98688 (1), Q992532 (1), Q993831 (1), and (blank) (96). The 'Facet by choice counts' option is selected. On the right, the main table view shows 96 matching rows out of 426 total. A context menu is open over the table, with 'Remove all matching rows' highlighted in yellow. The table columns are 'Inventory number' and 'Bestandesbildner'. The table shows rows with Q-numbers and names, such as '15-0: ALS Arès, Georges' and '15-0: ALS- L'Andelyn, Charles de'.

Then import the exported file with a new name. Filter it the same way, choose blank, invert the choice and remove the rows.

The screenshot shows the OpenRefine interface. The 'Facet / Filter' panel is active, showing the same text facet on the 'Q.-ID' column. The facet has 329 choices, sorted by name, with counts. The choices include Q76483 (-1), Q76815 (-1), Q77109 (-1), Q813089 (-1), Q824840 (-1), Q87292 (-1), Q98688 (-1), Q992532 (-1), Q993831 (-1), and (blank) (-96). The 'Facet by choice counts' option is selected. The 'invert' option is checked, and the 'include' button is visible.

3. Processing the data with Q-numbers

3.1. Import existing "archives at" (P485) values

Other entries with an "archives at" may already exist in Wikidata items:

"Edit column" -> "Add columns from reconciled values" -> add property "archives at" (P485)
(cp. 1.3)

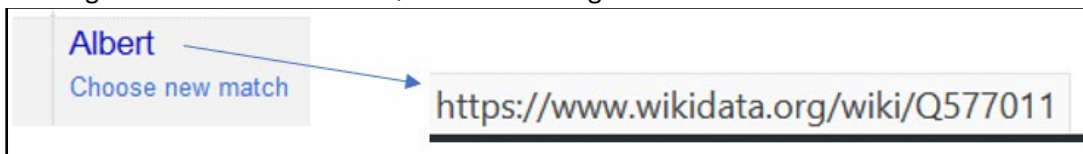
3.2. Existing P485

If there are already "archives at"-entries from your institution (referring to the same holdings): These entries will not be overwritten if their content is not exactly the same. So best make sure that you only add new links or content.

If you have existing Wikidata items you simply want to import to Wikidata see below chapter 6 "Import to Wikidata".

4. Processing new items (= dataset without Q-numbers)

For a higher data quality enrich your data with further information: e.g. start a reconciliation of the given name and add the Q-number to the given names where available



It is also advisable to add the gender of the person:

The first option is to match the gender of the given name (male/female given name): given names in Wikidata have an "instance of"-property, so you can add the value after reconciling the given names



If there is more than one given name, it might be necessary to split these into separate cells:

"Edit column" -> "Split into several columns...", you can use a regular expression as a separator (e.g. "\s" for a blank)

Split column recon.value(label) into several columns

How to Split Column

by separator

Separator regular expression

Split into columns at most (leave blank for no limit)

by field lengths

After Splitting

Guess cell type

Remove this column

Another option is to reconcile your data against the GND and add gender if available (see chapter 5).

5. GND

5.1. Data preparation for reconciliation against GND

Add name (=preferred name) in GND-compliant structure:
 [Family name][,][given name] (for example: *Alpari, Julius*)

5.2. Reconciling against GND

“Reconcile” -> “Start reconciling” -> Services: GND

This service has to be added first with “Add Standard Service...”:

<https://lobid.org/gnd/reconcile>

See above: add also a new column with the name to reconcile and then start reconciling against GND using details as date of birth etc.

Reconcile each cell to an entity of one of these types:

- Normdatenressource
AuthorityResource
- Individualisierte Person
DifferentiatedPerson
- Körperschaft
CorporateBody

Also use relevant details from other columns:

Column	Include? As Property	
Signatur	<input type="checkbox"/>	
Label	<input type="checkbox"/>	
Q ID	<input type="checkbox"/>	
family name (Q101352)	<input type="checkbox"/>	
given name (Q202444)	<input type="checkbox"/>	
Date of birth	<input checked="" type="checkbox"/>	gebu

Select an item from the list:

Geburtsdatum	dateOfBirth
Geburts- und Sterbedatum	dateOfBirthAndDeath
Geburtsort	placeOfBirth
Geburtsort (Literal)	placeOfBirthAsLiteral

In Wikidata you need a description of your items. The GND is a good resource to get the German description: “add columns from reconciled values” and harvest biographical data or other information from GND.

Add Property	Preview																						
<input type="text"/> Suggested Properties <ul style="list-style-type: none"> Adelstitel Adelstitel (Literal) Affiliation Affiliation (Literal) Akademischer Grad Alte Normnummer Beruf oder Beschäftigung Beruf oder Beschäftigung (Literal) Berufliche Beziehung Bevorzugter Name Beziehung, Bekanntschaft, Freundschaft Biografische oder historische Angaben Charakteristischer Ort 	<table border="1"> <thead> <tr> <th>Name_gnd_rec</th> <th>Biografische oder historische Angaben remove configure</th> </tr> </thead> <tbody> <tr> <td>Jaccard, Louis</td> <td>Spieldosenhersteller in der Firma Jaccard & Co. Waadtländer Grossrat, Regierungsratthalt Sainte-Croix</td> </tr> <tr> <td>Reiser, Johann</td> <td>Johann Reiser führte ab 1859 ein Atelier in welchem 1874 von Gottlieb Abderhalden übergeben wurde. Johann Reiser war danach in Kappel am Albis Fotograf tätig.</td> </tr> <tr> <td>Fengler, Bruno</td> <td></td> </tr> <tr> <td>Gravier, Gabriel</td> <td></td> </tr> <tr> <td>Krayer, Adolf</td> <td>Seidenfabrikant/-händler</td> </tr> <tr> <td>Keller, Salomon Heinrich</td> <td></td> </tr> <tr> <td>Keller, Salomon Heinrich</td> <td></td> </tr> <tr> <td>Keller, Salomon Heinrich</td> <td></td> </tr> <tr> <td>Rudolf, Urs Josef</td> <td></td> </tr> <tr> <td>Mieusement, Séraphin Médéric</td> <td></td> </tr> </tbody> </table>	Name_gnd_rec	Biografische oder historische Angaben remove configure	Jaccard, Louis	Spieldosenhersteller in der Firma Jaccard & Co. Waadtländer Grossrat, Regierungsratthalt Sainte-Croix	Reiser, Johann	Johann Reiser führte ab 1859 ein Atelier in welchem 1874 von Gottlieb Abderhalden übergeben wurde. Johann Reiser war danach in Kappel am Albis Fotograf tätig.	Fengler, Bruno		Gravier, Gabriel		Krayer, Adolf	Seidenfabrikant/-händler	Keller, Salomon Heinrich		Keller, Salomon Heinrich		Keller, Salomon Heinrich		Rudolf, Urs Josef		Mieusement, Séraphin Médéric	
Name_gnd_rec	Biografische oder historische Angaben remove configure																						
Jaccard, Louis	Spieldosenhersteller in der Firma Jaccard & Co. Waadtländer Grossrat, Regierungsratthalt Sainte-Croix																						
Reiser, Johann	Johann Reiser führte ab 1859 ein Atelier in welchem 1874 von Gottlieb Abderhalden übergeben wurde. Johann Reiser war danach in Kappel am Albis Fotograf tätig.																						
Fengler, Bruno																							
Gravier, Gabriel																							
Krayer, Adolf	Seidenfabrikant/-händler																						
Keller, Salomon Heinrich																							
Keller, Salomon Heinrich																							
Keller, Salomon Heinrich																							
Rudolf, Urs Josef																							
Mieusement, Séraphin Médéric																							

5.3. Add GND ID

“Edit column” -> “Add column based on column”: `cell.recon.match.id`

Add column based on column Name_gnd_rec

New column name

core-views/addasdasd set to blank store error copy value from original column

Expression Language No syntax error

6. Import to Wikidata

6.1. Enrich existing Wikidata items via OpenRefine

Import the missing data from the “Set of persons that already have an item in Wikidata” via Schema in OpenRefine.

244 records Schema Issues 217 Preview

The Wikidata schema below specifies how your tabular data will be transformed into Wikidata edits. You can drag and drop the column names below in most input boxes: for each row, edits will be values in these columns.

Nr.	Nachname	Vorname	GND-Ansetzungsform	GND-Nummer	GND_DateOfBirth	GND_dateOfDeath	Label	archives at2	archives at	WD_Nummer	geb.	WD_date of birth	Abgleich g

Labels:

Terms: no labels, descriptions or aliases added

Statements

archives at

▼ 1 references

GND ID

6.2. Enrich data in Wikidata with QuickStatements

QuickStatements allows to copy/paste data from a spreadsheet and to add mass data. Therefore you have to add double quotes in all columns with text strings, that you want to add to Wikidata e.g. the archival reference number or the URL.

Use edit cells - transform and the expression `"\""+value+"\""`

Custom text transform on column Inventory number

Expression Language No syntax error.

Preview History Starred Help

row	value	"\""+value+"\""
2	CH-000015-0: ALS Cendrars	"CH-000015-0: ALS Cendrars"

For import via QuickStatements you have to link any content to the correct Q-IDs or P-IDs manually instead of using a schema. Therefore you will have to add some more columns. See also <https://www.wikidata.org/wiki/Help:QuickStatements>

You can use OpenRefine or any other spreadsheet tool. Export the OpenRefine project in a spreadsheet format.


Example of a table, that is ready to copy/paste in QuickStatements

Label	Q-ID	P485	archive	P217	Inventor y number - string	S854	Link HelveticArchives	S813	Date
Markus Kägi	Q49437461	P485	Q675617	P217	"CH-000015-0: SLA-Kaegi"	S854	"https://www.helveticarchives.ch/detail.aspx?ID=931331"	S813	+2019-12-05T00:00:00Z/11
Otto Basler	Q2038184	P485	Q675617	P217	"CH-000015-0: SLA-Otto-Basler"	S854	"https://www.helveticarchives.ch/detail.aspx?ID=203198"	S813	+2019-12-05T00:00:00Z/11
Ilma Rakusa	Q124694	P485	Q675617	P217	"CH-000015-0: SLA-Rakusa"	S854	"https://www.helveticarchives.ch/detail.aspx?ID=124622"	S813	+2019-12-05T00:00:00Z/11
Jürg Schubiger	Q124576	P485	Q675617	P217	"CH-000015-0: SLA-Schubiger"	S854	"https://www.helveticarchives.ch/detail.aspx?ID=1057618"	S813	+2019-12-05T00:00:00Z/11

Label is only needed for reasons of quality assurance. Q-ID is the column from OpenRefine where you already added the Q-IDs of the persons with reconciling. P485 is the property “archives at” and as archive you add the Q-ID of the holding institution. P217 is the property for “inventory number” followed by the string you wanted to fill in there.

To add a reference you have to use S instead of P. P854 is the property for “reference URL” and P813 the property for “retrieved”. The date has to be written in this format. The timestamp should use the form T00:00:00 in order to avoid errors on import.

Copy/paste the spreadsheet table without titles in QuickStatements and transform with the “Import V1 commands” button.

QuickStatements English  [New batch](#) | [Last batches](#) | [Chat](#) | [Git](#) | [Help](#)

Create new command batch for Wikidata as batch name

Q49437461	P485	Q675617	P217	"CH-000015-0: SLA-Kaegi"	S854	"https://www.helveticaarchives.ch/detail.aspx?ID=931331"	S813	+2019-12-05T00:00:00Z/11
Q2038184	P485	Q675617	P217	"CH-000015-0: SLA-Otto-Basler"	S854	"https://www.helveticaarchives.ch/detail.aspx?ID=203198"	S813	+2019-12-05T00:00:00Z/11
Q124694	P485	Q675617	P217	"CH-000015-0: SLA-Rakusa"	S854	"https://www.helveticaarchives.ch/detail.aspx?ID=1246622"	S813	+2019-12-05T00:00:00Z/11
Q124576	P485	Q675617	P217	"CH-000015-0: SLA-Schubiger"	S854	"https://www.helveticaarchives.ch/detail.aspx?ID=1057618"	S813	+2019-12-05T00:00:00Z/11

[Import V1 commands](#) [Import CSV commands](#)

Status: 0% (0) of 1.

1	init	Markus Kägi (Q49437461)	ADD	Statement	archives at [P485]: Swiss Literary Archives [Q675617]			
2	init	Markus Kägi (Q49437461)	ADD	Qualifier to	archives at [P485]: Swiss Literary Archives [Q675617]	inventory number [P217]: "CH-000015-0: SLA-Kaegi"		
3	init	Markus Kägi (Q49437461)	ADD	Sources to	archives at [P485]: Swiss Literary Archives [Q675617]	reference URL : "https://www.helveticaarchives.ch/detail.aspx?ID=931331"		retrieved [P813]:2019-12-05
4	init	Otto Basler (Q2038184)	ADD	Statement	archives at [P485]: Swiss Literary Archives [Q675617]			
5	init	Otto Basler (Q2038184)	ADD	Qualifier to	archives at [P485]: Swiss Literary Archives [Q675617]	inventory number [P217]: "CH-000015-0: SLA-Otto-Basler"		
6	init	Otto Basler (Q2038184)	ADD	Sources to	archives at [P485]: Swiss Literary Archives [Q675617]	reference URL : "https://www.helveticaarchives.ch/detail.aspx?ID=203198"		retrieved [P813]:2019-12-05
7	init	Ilma Rakusa (Q124694)	ADD	Statement	archives at [P485]: Swiss Literary Archives [Q675617]			
8	init	Ilma Rakusa (Q124694)	ADD	Qualifier to	archives at [P485]: Swiss Literary Archives [Q675617]	inventory number [P217]: "CH-000015-0: SLA-Rakusa"		
9	init	Ilma Rakusa (Q124694)	ADD	Sources to	archives at [P485]: Swiss Literary Archives [Q675617]	reference URL : {"type": "unknown", "text": "https://www.helveticaarchives.ch/detail.aspx?ID=1246622" }		

You can now check if the data looks complete and correct, and then run the import. Progress and errors will be reported. With this method you can update already existing statements with qualifiers and sources and avoid creating duplicate statements for the same archival funds.

6.3. Add new Wikidata items via OpenRefine

For persons who have no item in Wikidata a new item has to be generated

187 records Schema Issues 537 Preview

The Wikidata schema below specifies how your tabular data will be transformed into Wikidata edits. You can drag and drop the column names below in most input boxes: values in these columns.

Nr.	Nachname	Vorname	GivennameWD	instance of 1	GND-Ansetzungsform	Biografische oder historische Angaben2	GNDnumber	GND_DateOfBirth	GND_dat
	Abgleich geb.WD	gest.	WD_date of death	Abgleich_gest.WD	Anmerkungen	Signatur	Link		

Label remove

Terms

Label	en	Label	<input type="checkbox"/> override if present	remove
Label	de	Label	<input type="checkbox"/> override if present	remove
Description	de	Biografische oder historische A...	<input type="checkbox"/> override if present	remove
Alias	en	Anmerkungen	<input checked="" type="checkbox"/> override if present	remove
Alias	de	Anmerkungen	<input checked="" type="checkbox"/> override if present	remove

+ add term

Statements

instance of remove

human remove

[0 references](#)

+ add qualifier
+ add reference
+ add value

date of birth remove

geb. remove

[1 references](#)

copy remove

stated in Integrated Authority remove

+ add
+ add reference
+ add value

date of death remove

gest. remove

[1 references](#)

copy remove

stated in Integrated Authority remove

+ add
+ add reference
+ add value

archives at remove

Zentralbibliothek Zürich remove

inventory num Signatur remove

[1 references](#)

copy remove

reference URL Link remove

retrieved 2019-10 remove

+ add
+ add reference
+ add value

given name remove

Givenname... remove

[0 references](#)

+ add qualifier
+ add reference
+ add value

inventory num	Signatur	remove
		+ add qualifier
▼ 1 references		
copy		
reference URL	Link	remove
retrieved	2019-10	remove
		+ add
		+ add reference
		+ add value
given name	Givenname...	remove
		+ add qualifier
		► 0 references
		+ add reference
		+ add value
GND ID	GNDnum...	remove
		+ add qualifier
		► 0 references
		+ add reference
		+ add value
sex or gender	instance o...	remove
		+ add qualifier
		► 0 references
		+ add reference
		+ add value

Upload process: “Extensions: Wikidata” -> “Upload edits to Wikidata”
 Check your results and the correct import of the data carefully.

Tips and tricks

- Be careful with diacritics etc.: some characters cause data import problems (check the number of generated rows when creating a new project).
- If you have more than one row per record in your original file, change the view under “Show as” from default “rows” to “records”
- Take the structure of the data into account. If the result of the matches are insufficient, it is helpful to adjust the data structure (e.g. when reconciling against the GND, remove all hyphens from the names first and replace them with a blank space).
- Reconcile against existing Wikidata items shortly before import because Wikidata grows quickly and you want to avoid duplicates.
- If you have any problems with some records (e.g. divergent dates or already existing duplicates in Wikidata), you better separate these records. Open a new project for these problematic cases and check this list in a separate step (manually if required).

Additional resources (last retrieved: January 16 2020)

<https://github.com/OpenRefine>

<https://histhub.ch/reconciling/> (German)

<http://blog.lobid.org/2018/08/27/openrefine.html>

<https://github.com/OpenRefine/OpenRefine/wiki/Understanding-Regular-Expressions>

<https://docs.oracle.com/javase/tutorial/essential/regex/>

<https://www.regular-expressions.info/>